www.ijprems.com
editor@ijprems.com

INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)
Vol. 02, Issue 05, May 2022, pp : 349-352

e-ISSN : 2583-1062
Impact Factor : 2.205

USING MACHINE LEARNING FOR SENTIMENT ANALYSIS THROUGH METICULOUS APPROACH

Sakshi Dongarwar[*1], Komal Katwe[*2], Prof. Vanita Buradkar[*3]

[*1,2,3]CSE Dept., RCERT, Chandrapur, India.

ABSTRACT

Sentiment analysis, also known as opinion mining, is an artificial intelligence approach that uses machine learning and deep learning libraries and models to assess and predict human emotions in a given text, which might be positive, negative, or neutral.[5] Most academics and business models employ this method for text mining and understanding human attitudes, and it is one of the most popular active study areas in artificial intelligence. It's difficult to manually assess a big number of data sets, which is where the sentiment analysis model comes in. Movie reviews may be utilised as a dataset for training and preprocessing. [4].

1. INTRODUCTION

Sentiment refers to a person's expression, emotions, attitude, or any other viewpoint.[1] There are three basic categories of sentiment: negative, neutral, and positive. Everyone nowadays utilises the internet or social media to share their views, opinions, and feelings about a certain issue.[3]The internet has evolved into a collection of viewpoints that must be studied in order to comprehend them. a person's criticisms, ideas, or sentiments Many social networking sites, business models, and product managers utilise feedback forms to collect useful data from users and then use it to improve their products. analyse it in order to have a better understanding of their input on that specific issue and, as a result, enhance their business As a result, adjust your methods or job efficiency.[6]

Sentiment analysis is the act of deciphering what is stated in a text, detecting the user's viewpoint, and categorising it as positive, negative, or neutral. Sentiment analysis is a helpful tool for both consumers and producers in today's society. Microblogs, social networking sites, forums, and communities are used by a vast number of individuals.It may be used to express one's thoughts and judgments about a topic or a product on the internet, or it can be used to convey one's opinions and judgments about a topic or a product You can be anything. Sentiment analysis is used by data analysts, business analysts, and many others to swiftly analyse data. Analyze the information and feedback you've gathered. A dataset of movie reviews from the United States is used in this study.[7]

The model is trained and tested using Amazon website.This collection of data includes reviews of several mobile phones having a polarity a number of reviews There are primarily two classifiers: Naïve On the dataset, Bayes and logistic regression are used to predict positive and negative sentiment. a user's input text or a review [1]

2. MACHINE LEARNING

Machine Learning is a branch of artificial intelligence that functions similarly to a human brain in that it learns and recognises objects on its own. It employs a dataset to train the model, as well as an algorithm to assess and enhance the model's learning efficiency. and the creation of decision-making patterns[2] Various classification algorithms were evaluated with the supplied training data in order to determine the optimal method for the model. Because of the overfitting of the model with the training dataset, the same training dataset is not utilised for measuring the efficiency of the methods.[5]

The outcomes will be inaccurate. It is advised to keep a safe distance from this problem. Sort the data into two groups: training and testing. There are a few methods for separating a pair of shoes. The dataset is divided into two parts: a test and a training set. Initially, the primary goal was to employ 60% of the dataset.as a train set, with the remainder of the set being used for testing. There is, nevertheless, a risk of overfitting. until the test set is complete.

Using machine learning algorithm for sentiment analysis

Supervised learning algorithms

It is one of the most extensively used machine learning algorithms that trains the data and develops the model for testing using a well-labeled data set. The labelled data set is a dataset that contains both input and output parameters. The model picks up information from the train set.

It takes a data collection and extracts knowledge or logic from it, after which the input is supplied into the model. It then combines the input with the logic to produce the result.

3. LITERATURE REVIEW

One of the first methods for analysing the opinion or mood of a document is to use a lexicon [3]. It extracts feelings from the text, compares it to its vocabulary, and assigns a strength score to the text using an external dictionary or a

www.ijprems.com
editor@ijprems.com

INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)
Vol. 02, Issue 05, May 2022, pp : 349-352

e-ISSN : 2583-1062
Impact Factor : 2.205

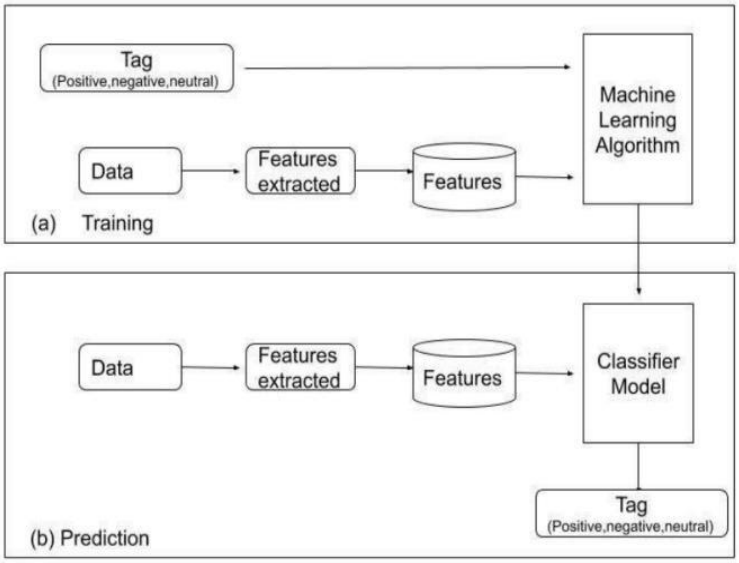
set of bags. The data is divided into three categories based on the determined score: positive, negative, and neutral. The score in terms of semantics

The polarity and strength of a particular word are calculated using the calculator, commonly known as So-cal. So-cal gives it a score in the lexicon. This strategy is simple to comprehend and implement. Manually adjustable, although there are several drawbacks to this method, one of which is that it is not scalable. One of the method's limitations is that it is unable to discern between different types of data.

Another technique, named Latent Aspect Rating Analysis on Review Text Data: A Rating Regression Approach [5], was developed by

Hongning Wang, Lu Yue, and Chengxiang Zhai. Which first classifies the data using a rating approach, then extracts the appropriate text to estimate its polarity using a text extraction method.as well as emotion Many online shopping companies employ this strategy in conjunction with the usage of social media.The user's feedback and reviews are analysed using a neural network algorithm. The foundationThe disadvantage of this technique is that user reviews are confined to a few specific features. a few pre-defined terms to describe the feature

The most popular method is machine learning, which involves training a model with a set of algorithms [6]. The dataset is divided into two sets: the train set and the test set, so that the model can be overfit to a specific dataset while being tested. For the time being, it is not feasible for the The raw dataset is first cleaned and then the algorithm is used to train the model on it. The tf-idf factor was used to vectorize the data. Sentiment analysis on raw data using machine learning is powerful tool. Challenging procedure owing to a variety of factors, including the model's inability to detect sarcasm. the content, as well as the positive and negative emotion included inside the same text data These issues have been resolved.



4. METHODOLOGY

Basic architecture of how sentiment analysis work is shown in figure Architecture of Sentiment Analysis of Proposed architecture has 3 steps –

a) Data cleaning b) Vectorization c) Training a machine learning Model

a) Data cleaning

Prior to vectorization, the data must be cleaned, which means there must be no punctuation or stopwords that do not contain any relevant information since they cause the machine learning model to perform incorrectly. There is a python library called regular expression or regular expressions that may be used for purpose..Page breaks, punctuation, and other special characters may all be removed using the re library.Only after deleting all extraneous data and stopwords is the cleaned processed data returned to the user.acquired.

b) Vectorization Now that the data has been cleaned, the data must be vectorized since, with such a vast number of datasets, a complex function might become a major stumbling block in the model, and it will take a long time to run an algorithm. To minimise the model's temporal complexity and turn a text into Vectorization is employed in the numerical 2-d matrix form. The best way to undertake sentiment analysis is to The tf-idf vectorizer is a vectorization algorithm that may be used in vectorization.

www.ijprems.com
editor@ijprems.com

INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)
Vol. 02, Issue 05, May 2022, pp : 349-352

e-ISSN : 2583-1062
Impact Factor : 2.205

c) Training a machine learning model

To train a robust ML model, you'll need a labelled dataset. The machine learning model will learn various patterns in the data and will be able to predict sentiment for unseen text.

The following procedures must be followed in order to train a bespoke sentiment analysis model:

For sentiment analysis,

1.collect a raw labelled dataset.

2.Text preprocessing

3.Numerical text encoding

4.Selecting the best machine learning algorithm

5.ML model hypertuning and training

6.Prediction [8]

5. RESULTS AND DISCUSSION

Score calculated using Essemble Classification model, Logistic Regression, Multinomial BernoulliNB , Decision Tree, knn & SVM are as follow

LogisticRegression 0.948250

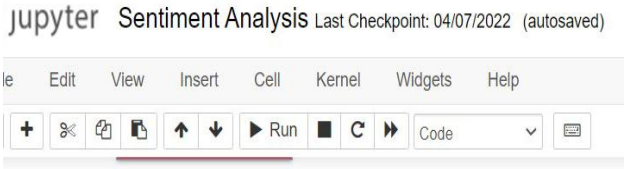
Multinomial 0.922625

\nBernoulliNB 0.813250

DecisionTree 0.940125

EssembleClasification 0.964750

SVM 0.943375 knn 0.934625

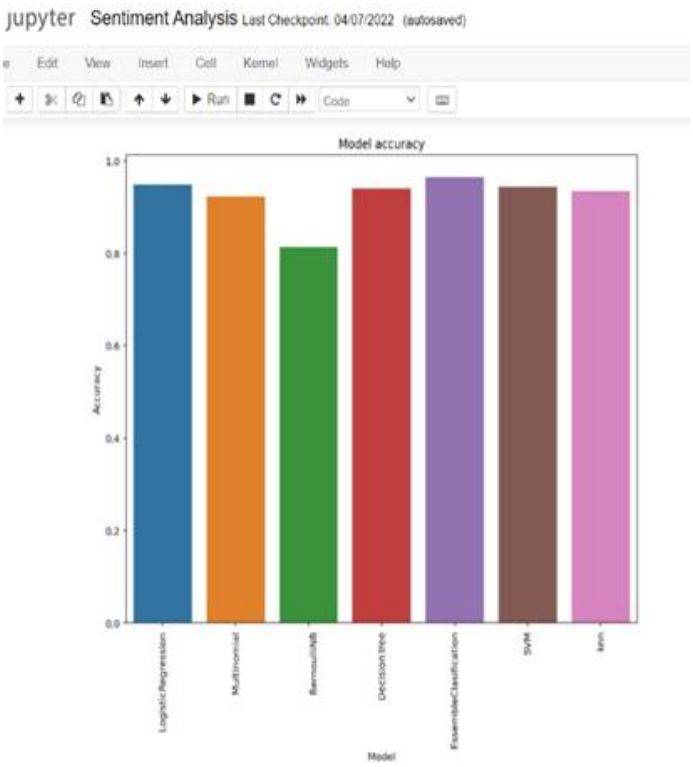



```
Please wait. This may take a few minutes

Analyzing Logistic Regression

Analyzing Multinomial NB

Bernoulli NB

Analyzing Decision Tree

Analyzing Ensemble Classifier

Analyzing SVM

Analyzing k-NN


Calculating Accuracy of each model.


                  Model  Accuracy
0    LogisticRegression  0.948250
1           Multinomial  0.922625
2         \nBernoulliNB  0.813250
3          DecisionTree  0.940125
4  EssembleClasification  0.964750
5                   SVM  0.943375
6                   knn  0.934625
```

www.ijprems.com
editor@ijprems.com

INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)
Vol. 02, Issue 05, May 2022, pp : 349-352

e-ISSN : 2583-1062
Impact Factor : 2.205



Efficiency of all the models are calculated as shown in figure

This shows that while training the model with the reviews data set Essemble Clasification model is much better than Logistic Regression, Multinomial BernoulliNB , DecisionTree, knn & SVM.

6. CONCLUSION

Sentiment analysis is a prominent issue in machine learning research, and there are a variety of machine learning techniques for training and testing a model. The sentiment analysis of movie reviews as a data set is demonstrated in this paper utilising algorithms. Logistic Bayes vs naive Bayes regression as a baseline for training the model, and the relationship between language models and the Naive Bayes model delivers better outcomes Machine learning is a popular technique for sentiment analysis, however the model is still unable to interpret sarcastic language or the difference between negative and positive sentiment.

In the same text, there is a similar emotion. The model may also be trained on a variety of data sources. It may also be used for various types of social media sentiment analysis. The project will continue in the future.

7. REFERENCES

[1] Cambria, E., Schuller, B., Xia, Y., and Havasi, C., "New avenues in opinion mining and sentiment analysis," IEEE Intell. Syst., 2013, doi:10.1109/MIS.2013.30.

[2] Collomb, A., Costea, C., Joyeux, D., Hasan, O., and Brunie, L., "A Study and Comparison of Sentiment Analysis Methods for Reputation Evaluation," 2013.

[3] Taboada, M.,Brooke, J., Tofiloski, M., Voll, K., and Stede, M., "Lexiconbasedmethods for sentiment analysis," Comput. Linguist., 2011, doi:10.1162/COLI_a_00049.

[4] Medhat, W., Hassan, A., and Korashy, H., "Sentiment analysis algorithms and applications: A survey," Ain Shams Eng. J., 2014, doi: 10.1016/j.asej.2014.04.011.

[5] Wang, H., Lu, Y., and Zhai, C., "Latent aspect rating analysis on review text data: A rating regression approach," Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ISBN 9781450300551, 2010, doi:10.1145/1835804.1835903.

[6] Boiy, E. and Moens, M.F., "A machine learning approach to sentiment analysis in multilingual web texts," Inf. Retr. Boston., 2009, doi:10.1007/s10791-008-9070-z.

[7] Manning, C.D., Raghavan, P., Schutze, H., Manning, C.D., Raghavan, P., and Schutze, H., "Scoring, term weighting, and the vector space model," Introduction to Information Retrieval, 2012, doi:10.1017/cbo9780511809071.007.

[8] https://towardsdatascience.com/5-ways-to-develop-a-sentiment-analyser-in-machine-learning-e8352872118